DOCUMENT RESUME

ED 155 219 TM 007 365

AUTHOR Tittle, Carol Kehr
TITLE Evaluation and Decision Making: Developing a Method to Link Program Funding Decisions and Outcome Evaluation.
PUB DATE [Mar 78]
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Budgeting; *Decision Making; *Educational Assessment; *Evaluation Criteria; Evaluation Methods; Evaluators; *Grants; Predictor Variables; Program Administration; *Program Evaluation; Research Problems; *Research Utilization; State Departments of Education; State Programs; Summative Evaluation
IDENTIFIERS Vocational Education Act 1976

ABSTRACT

There is a continuing need in evaluation research for the establishment of a relationship between evaluation findings and decision making. A method is proposed for a particular situation: annual funding decisions for projects in a large grant program in vocational education. Outcome and predictive impact variables were ranked by three groups of decision makers on a pilot study. The groups included the Director of the State Department of Education division responsible for funding decisions, the supervisors who make funding decisions, and the supervisors from related bureaus who review and contribute to the decision making process. Statements concerning the impact of vocational education programs on students, employers, and the State Department of Education--to be used as program evaluation criteria--were sorted into twelve outcome impact and nine predictive impact statements. Each statement was ranked and rated for importance by the decision makers. Results showed high agreement on the ranking and rating of outcome impact statements, and discrepancies on the predictive impact statements. A validation study has been designed. Evaluators can assist decision makers in identifying important outcomes; and in the process, define the decision to be made, the time when it is made, and the data required to link evaluation and decision making. (Author/JAG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Evaluation and Decision Making: Developing a Method to Link Program Funding Decisions and Outcome Evaluation[1]

Carol Kehr Tittle

Institute for Research and Development in Occupational Education
Center for Advanced Study in Education
Graduate School and University Center
City University of New York

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Carol Kehr Tittle

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED155219

Introduction

A continuing problem in evaluation research is establishing a relationship between evaluation findings and decision making. The objective of the work described here is to improve this relationship in one setting for decision making: annual funding decisions for grant programs under the Vocational Education Act of 1976. The method is being developed in a pilot project with a state education department (SED) where the funding decisions are made.

The specific purposes of the pilot project are: 1. to develop definitions of outcome variables that will, when combined, identify a "high" impact project, that is, a project which is high on important outcome dimensions for students, employers, or the granting agency; and 2. to relate these definitions to funding decisions. For purpose 2., definitions of impact variables (evaluation outcome variables) are translated into "predictive impact" variables that can be determined for each proposal at the time of the application for funding, thus providing some "objectively determined" data to be combined with the other information entering the funding decision making process. The development of this method can improve the relationship between evaluation and decision making by identifying the most important outcomes for program evaluation and decision making for both project directors and SED decision

[1] Paper prepared for presentation at the annual meeting of the American Educational Research Association, Toronto, March 1977.

Session D-32

TM007 365

Printed in U. S. A.

makers. By focusing attention on the aspects of these variables that can be known at the time of funding, it is proposed that the likelihood of funding "high impact" projects will be increased.

Related Research

Davis and Salasin (1975) have summarized many of the issues in the use of evaluation results, including statements by evaluators that their results are not used and those by administrators that evaluation findings are not available when decisions have to be made. Although there is much discussion of the need to relate evaluation and decision making, there have been few efforts to specify the manner in which this might occur.

Hemphill (1969) provided an early example of the use of educational evaluation data in a formal decision theory model. He illustrated several uses of decision theory in evaluation. In one instance, an individual decision maker decided whether to install a new nursery school, and also examined whether to carry out the evaluation study, again within a decision theory framework. More recently, Edwards, Guttentag, and Snapper (1975) have proposed and applied a method called multi-attribute utility measurement (MAUM) to assist the Office of Child Development in defining the major dimensions of importance in developing priorities for funding research projects. The multi-attribute utility measurement method is one of a set of methods classified as decision aids by Slovic, Fischoff, and Lichtenstein (1977), as opposed to formal behavioral decision theory models.

Another decision-aid model, social judgment theory (SJT), has been described by Hammond and his colleagues, and developed into an interactive computer program. Both MAUM and SJT are well-described in the literature (Edwards, 1976; Gardiner and Edwards, 1975; Guttentag, and Snapper, 1975; Guttentag

1973; Guttentag and Snapper, 1974; Hammond and Adelman, 1976; Hammond, Stewart, Adelman, and Wascoe, 1975; Hammond, Stewart, Brehmer and Steinmann, 1975). There is considerable overlap in the basic ideas of the multi-attribute utility measurement method and the social judgment theory as it is applied by Hammond. One main difference, however, is in the nature of the judgment task presented to decision makers in order to determine the most important dimensions or variables for decision making and their relative importance (say, for example, the major outcome variables of a program or the priorities among a set of goals for a funding program). The social judgment theorist, following Brunswik and the importance of representational validity, presents combinations of variables as they would occur in real projects in the decision setting in order to elicit the utilities or values placed on the major variables by the decision maker. The variables are judged (weighted) independently in the MAUM as developed by Edwards.

The literature on decision aids is one area of relevant research; the second area consists of defining "impact" for evaluation purposes. Papers concerned with the analysis of impact and impact assessment methodology have not always dealt with the problem of how to define impact. However, representative definitions include, "the capacity of a program to cause changes in those who are exposed to it" (Houston, 1972), and "the difference between what happens with the intervention and what would happen without" (Levine, 1976). Bernstein and Freeman(1975) deal with the definition problem by presenting the requirements for impact measurement:

1. document the extent to which the social action program has or has not achieved its stated goals;

2. attribute any effects or changes that are discovered to the implementation of the program;
3. delineate, if possible, the conditions or combinations of conditions under which the program is most effective;
4. delineate, if possible, any unanticipated consequences or side effects of the program.

This definition assumes a common set of goals across programs, if one objective of evaluation is to permit a comparison of the impact of different programs within a major funding program. Sirois and Iwanicki (1977) have noted the between-program comparison problem, where program goals lack specificity and regularity. The present project is concerned with a somewhat different perspective, since not all programs funded under the VEA can be expected to meet all the goals or priorities of the legislation.

Hu and Stromsdorfer (1975) defined general criteria for measuring the educational and economic impact of research and demonstration projects in vocational education. Two types of impact of vocational education were identified: 1. intermediate impact or goals, and 2. final output or ultimate goals. The first type included: modification or revision of curricula; re-allocation of funds within the educational system; effects on students' aptitudes and school performance; number of graduates produced; percentage of graduates working in occupations for which they were prepared; improvement in student attendance; and sense of fulfillment in vocational education teachers after developing a new program. The final output included labor market performance of students (wages, employment, job satisfaction) and educational attainment.

Hu and Stromsdorfer's list can be viewed as a general set of criteria. In addition to the types of variables in their list, the VEA for 1976

provides that training for special populations is also an important goal for VEA programs. The diversity of (legitimate) outcomes means that not all programs will have the same set of objectives. And, there is a concern expressed by state decision makers and local program administrators that not every program can meet the same set of standards. When programs serve diverse populations, as in the vocational education legislation, it is more realistic to have a set of "impact" variables to evaluate program impact, not all of which are expected to apply to each program. Projects can be judged to have a high impact by meeting some standards (that is, being high on some impact scales) but not on all. The same conclusion can be drawn for the predictive impact variables being identified in the current project. Where federal legislation has multiple goals and groups to be served, a set of important outcomes that are operationally defined may permit identification of "high impact" projects and also permit local needs to be met. Yet, the definition of the the impact variables and their use in funding decision making may serve to focus local projects and evaluations on these same outcomes.

Methods and Results to Date

The work to date has generally followed the Edwards, Guttentag, and Snapper (1975) procedure to develop the impact dimensions. As noted by both Hammond, Stewart, Brehemer and Steinmann(1975) and Edwards (1976) the most important step is the first one; developing a clear understanding of the decision making process and developing the lists of variables to be considered for rating. Interviews were conducted with SED decision makers to develop a flow chart of the decision process for funding, and to elicit statements defining or critical to "high impact" projects. In addition, a review of the literature in vocational education was conducted to identify

other goals and objectives. An initial list of 104 statements related to impact of vocational education programs on students, employers, and the SED were sorted and reduced to 12 outcome (impact) statements. These statements were rephrased to nine predictive impact statements by identifying the variables that were known conditions to achieving the outcomes. (See Table 1 for sample statements.) Rank ordering and ratings of importance by decision makers were then obtained. At the time of the rating, all the dimensions were specified operationally, so raters had an idea of what eventual scale definitions might be, even though on a tentative basis.

Three rater "groups" were used. The first "group" consisted of the Director of the SED division responsible for funding decisions; the second group was the set of supervisors who make funding decisions; and the third group was a set of supervisors from related bureaus who also review and have a part in the decision making process. Agreement among the three "groups" of raters was measured by Kendall's Coefficient of Concordance ($\underline{W}$). The coefficients for the twelve outcome statements were .81 for the ranks and .94 for the ratings ($\underline{p} < .01$). Agreement was not as high for the predictive impact statements: $\underline{W} = .81$ ($.05 < \underline{p} < .01$) for the ranks; and $\underline{W} = .41$ ($.50 < \underline{p} < .30$) for the ratings of the importance of the predictive impact statements.

The high agreement among the raters on the set of ratings and rankings of the <u>outcome</u> impact statements was encouraging. It was not clear why there was a discrepancy between the agreement on ranks and that for the ratings for the <u>predictive</u> impact statements. As a result, there has been a revision of the two sets of statements, and a second set of ratings and rankings will be obtained. Some statements which were not ranked highly in either set of statements have been removed, and the two sets are now parallel.

Table 1
Sample Outcome Impact
Statements and Categories*

TRAINING OBJECTIVES ARE MET WITH MINIMAL COST PER STUDENT.

Training cost per student:

$500 or less ___ $501 to $1000 ___ $1001 to $1500 ___ $1501 or more ___

PROGRAM GRADUATES ARE WORKING IN OCCUPATIONS FOR WHICH THEY WERE TRAINED.

Percentage of program graduates employed in occupations for which trained (within first six months):

0-25% ___ 26-50% ___ 51-75% ___ 76-100% ___

THE VOCATIONAL EDUCATION NEEDS OF SPECIAL GROUPS ARE MET - THE ECONOMICALLY DISADVANTAGED, THE HANDICAPPED, AND PERSONS WITH LIMITED ENGLISH-SPEAKING ABILITY.

Percentage of students trained who are from these special groups:

0-25% ___ 26-50% ___ 51-75% ___ 76-100% ___

STUDENTS ARE TRAINED FOR OCCUPATIONS TRADITIONALLY DOMINATED BY THE OPPOSITE SEX.

Percentage of students in program who are trained for occupations traditionally dominated by the opposite sex.[a]

0-25% ___ 26-50% ___ 51-75% ___ 76-100% ___

[a] Occupations in which the proportion of women is less than 38%

EMPLOYERS ARE SATISFIED WITH GRADUATES OF THE PROGRAM

a. Percentage of graduates that employers rate as satisfactory on entry level skills:

   0-25% ___ 26-50% ___ 51-75% ___ 76-100% ___

b. Percentage of graduates that employers retain or promote (for a two-year period):

   0-25% ___ 26-50% ___ 51-75% ___ 76-100% ___

* Predictive Impact Statements are often the same, with the exception that they are stated in the future tense (...will be ...).

A summary of the method to date includes these steps:

1. interviews of decision makers and surveys of the literature to identify critical aspects of "high impact" projects;
2. Developing lists of outcome variables from 1., above;
3. "free" sorting of statements by evaluation staff;
4. For a reduced list (twelve or fewer statements), state both as outcome and predictive statements; develop operational definitions and sample scales for raters.
5. Obtain rankings and ratings of the preliminary set[1]

The ratings or value (utilities) attached to the dimensions can then be used in formal decision theory or a Bayesian decision theory approach (See Winkler, 1972).

The next steps in the project are to develop forms which can be used to provide the data needed for funding decisions and for evaluation (the predictive and outcome impact statements, respectively). These data provide the information necessary to change the sample categories given raters to categories based on distribution data. For example, the sample categories for cost per student in Table 1 are fictional. In order to know whether

---

[1] There are, obviously, any number of psychological scaling methods available, as well as the methods used in formal behavioral decision theory to obtain utilities (see Slovic, et al, 1977). The objective is to obtain a reduced set of statements and then to obtain weights for the final set of statements. However, it is not clear that other than unitary weights have great value. For example, Dawes (1973) and Hammond, Stewart, Adelman and Wascoe's studies (1975) had results suggesting that with multiple judges equal weights may result.

a project is high or low on cost per student, actual data must be obtained and grouped for comparable programs.

Data will be obtained for past projects on the outcome statements. A sample of evaluations will be judged as to their overall level of impact by decision makers and "scored" on the outcome variables by the evaluators. The relationship between the two measures will be obtained as one evidence of the "validity" of the impact dimensions. Also, ratings and rankings of the two sets of impact variables may be obtained from program directors at the Local Education Agencies (LEA's). These data will provide another perspective on the impact statements. Other sources of related validity data would be rankings of the variables by the State Advisory Council to the VEA. For the long term study of the validity of the predictive impact variables, there will need to be a follow-up rating of the overall level of impact for grant applications that were given predictive impact ratings.

There are limitations to the methods being proposed here. In the first place, much of the validation proposed is circular, as Slovic et al. (1977) have noted for the studies of the other methods of aiding decisions. Second, the impact variables described here, while clearly important for evaluation, are only one part of the information used for funding decisions. Other areas which are rated in funding include the general management plan for carrying out the project, the proposed staff, and the project evaluation plan. At this stage, it is not known what weight these variables carry in decision making, or whether they are over-riding variables. Unless standards are met in these other areas projects may not be funded, regardless of impact ratings.

Third, there are the political concerns of the method. One political aspect is the relationship between the evaluator and the administrative decision makers. The evaluator must be sensitive to staff needs in discussing and describing areas of work that are not typically accessible to outsiders. Particularly important is the need for detailed knowledge of the funding decision making process as it currently exists (and see Edwards, 1976, for similar caveats). A second political aspect is the relationship between decision makers and outside groups, such as state legislators and others with interest in the funding of local projects. From the evaluator's viewpoint, the development of well-defined criteria would appear to have benefits in providing a rationale for funding decisions. From a staff viewpoint, such a development may represent a loss of "degrees of freedom" in their decision making. Knowledge of the values or importance ratings given to the impact variables by other groups such as LEA project directors or local and state-level advisory councils might lessen this last concern.

Summary

The significance of the method being developed here is in providing one way in which evaluation and funding decisions can be linked in programs that are continuing and large scale. One characteristic of the programs funded under the VEA is that not every project will meet or be evaluated highly on all major impact dimensions. Evaluators can assist decision makers to establish the set of impact dimensions important for funding decisions. In the process, operational definitions of these variables are established for both evaluators and project directors. Evaluation and decision making can, in some cases, be more closely related if evaluators clarify the exact decisions to be made, who makes the decisions, timing of the decision, and

the nature of the data necessary to make the decision. In the VEA, new programs will not have available past evaluation data at the time of funding, but the use of predicive impact data can, it is hypothesized, increase the likelihood of funding projects that will later be judged as higher in impact.

In this example, and similar settings, evaluators may find that devoting effort to clarifying the types of decisions that can be made and the data that can be provided will increase the use of evaluation results over the long term.

References

Bernstein, I.N., & Freeman, H.E. Academic and entrepreneurial research. New York: Russell Sage Foundation, 1975.

Davis, H.R., and Salasin, S.E. The utilization of evaluation. In E.L. Struening and M. Guttentag (Eds.) Handbook of Evaluation Research, Vol. 1. Beverly Hills: Sage Publications, 1975.

Edwards, W. How to use multi-attribute utility measurement for social decision-making. Technical Report 001597 1-T. Los Angeles, Cal.: Social Science Research Institute, University of Southern California, 1976.

Edwards, W., Guttentag, M., & Snapper, K. A decision-theoretic approach to evaluation research. In E.L. Struening & M. Guttentag (Eds.) Handbook of evaluation research (Vol. 1). Beverly Hills: Sage Publications, 1975.

Gardiner, P.C., & Edwards, W. Public values: Multi-attribute utility measurement for social decision making. In M.F. Kaplan and S. Schwartz (Eds.) Human judgment and decision processes. New York: Academic Press, 1975.

Guttentag, M. Subjectivity and its use in evaluation research. Evaluation, 1973, 1, 60-65.

Guttentag, M., & Snapper, K. Plans, evaluations and decisions. Evaluation, 1974, 2, 58-64; 73-74.

Hammond, K.R., & Adelman, L. Science, values and human judgment. Science, 1976, 194, 389-396.

Hammond, K.R., Stewart, T.R., Adelman, L., & Wascoe, N. Report to the Denver city council and mayor regarding the choice of handgun ammunition for the Denver police department. Report No. 179. Boulder: University of Colorado, Institute of Behavioral Science, March 25, 1975.

Hammond, K.R., Stewart, T.R., Brehmer, B., & Steinmann, D.O. Social judgment theory. In M.F. Kaplan and S. Schwartz (Eds.) Human judgment and decision processes. New York: Academic Press, 1975.

Hemphill, J. K. The relationship between research and evaluation studies. In R. W. Tyler (Ed.), Educational evaluation: New roles, new means. (The 68th yearbook of the National Society for the Study of Education, Part II) Chicago: The University of Chicago Press, 1969.

Houston, T.R., Jr. The behavioral sciences impact-effectiveness model. In P.H. Rossi and W. Williams (Eds.) Evaluating social programs: Theory, practice, and politics. New York: Seminar Press, 1972.

Hu, Teh-Wei, & Stromsdorfer, E.W. An analysis of the impact of applied research and demonstration projects in vocational education. Office of the Assistant Secretary for Policy, Evaluation and Research, Department of Labor, July 1975.

Levine, A.S. Evaluating program effectiveness and efficiency: Rationale and description of research in progress. Welfare in Review, 1967, 5, 1-11.

Sirois, H.A., & Iwanicki, E.F. Delphi-discrepancy evaluation: A model for the quality control of federal, state, and locally mandated programs. Paper presented at the annual meeting of the American Educational Research Association, New York City, April 1977.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.

Winkler, R. L. An introduction to Bayesian inference and decision. New York: Holt, Rinehart, and Winston, 1972.